Intel® Omni-Path Software

Release Notes for V10.8.0.2

Rev. 1.0

January 2019

Contents

Tables

Intel® Omni-Path Software
Release Notes for V10.8.0.2   January 2019
4   Doc. No.: K48500, Rev.: 1.0

1.0   Overview of the Release

This document provides a brief overview of the changes introduced into the Intel® Omni-Path Software by this release. References to more detailed information are provided where necessary. The information contained in this document is intended as supplemental information only; it should be used in conjunction with the documentation provided for each component.

These Release Notes list the features supported in this software release, open issues, and issues that were resolved during release development.

1.1   *Important Information*

This release resolves the following issues (see Table 8 on page 15):

- STL-46606
- STL-46790
- STL-47546
- STL-47956
- STL-48661

1.2   Audience

The information provided in this document is intended for installers, software support engineers, service personnel, and system administrators.

1.3   Software License Agreement

This software is provided under license agreements and may contain third-party software under separate third-party licensing. Please refer to the license files provided with the software for specific details.

1.4   If You Need Help

Technical support for Intel® Omni-Path products is available 24 hours a day, 365 days a year. Please contact Intel Customer Support or visit http://www.intel.com/omnipath/support for additional detail.

1.5   Enhancements and New Features in this Release

The following enhancements and features are new in this release:

- New operating systems supported. See Table 2 on page 8 for details.
- New hardware supported. See Table 6 on page 10 for details.
- Support for multiple virtual fabric security.

- SE Linux PSM2 support with 13 labels.
- Support for libfabric OFI Version 1.6.
- Support for Sandia* OpenSHMEM 1.4.1.

1.6 Supported Features

- The list of supported operating systems is in Table 2 on page 8.
- The list of supported hardware is in Table 6 on page 10.
- Product Constraints described in Product Constraints on page 13.
- Active Optical Cables (AOC). For details, see the Cable Matrix at: http://www.intel.com/content/www/us/en/high-performance-computing-fabrics/omni-path-cables.html
- Support for the Enhanced Hypercube Routing Engine is outside the scope of Intel® OPA support. However, Intel partners may offer such support as part of their solutions. In addition there is an open source community who may be able to answer specific questions and provide guidance with respect to the Enhanced Hypercube Routing Engine.
- Support for Accelerated RDMA, also called Token ID (TID) RDMA, which is a Verbs protocol extension. See Accelerated RDMA Information on page 13 for details.
- Support for active optical cables (AOC) on server platforms using integrated HFI for OPA (commonly known as "-F").
- Support for GPUDirect* RDMA, which is a technology that enables a direct path for data exchange between a graphics processing unit (GPU) and a third-party peer device using standard features of PCI Express.
- Legacy BIOS Boot Mode Enhancements to support boot over fabric, custom board descriptions, and pre-boot platform configuration data for AOC support.
- Multi-endpoint functionality. See the *Intel® Performance Scaled Messaging 2 (PSM2) Programmer's Guide* for details.
- SNMP MIB support. See the *Intel® Omni-Path Fabric Switches Release Notes* for details.
- UEFI, TMM, and Firmware Tools delivered as standalone rpms.
- Support for Intel® C/C++ Compiler 18.0
- Intel® OPA support for cgroups.

1.7 Removed Features

Pallas Micro Benchmarks (PMB) has been removed in this release to be superseded by its new incarnation, Intel® MPI Benchmarks (IMB).

1.8 Release Packages

There are two Intel® Omni-Path Fabric Software packages:

- Basic for compute nodes
- IFS for management nodes

The Basic package includes:

- Software that installs the following packages to the distribution OpenFabrics Alliance* (OFA):

  — hfi1-firmware, libpsm2 (for all RHEL* and SLES* 12 SP2) and ibpsm2-2 (for SLES* 12 SP3 and newer), hfi1-diagtools-sw

  — Open MPI and MVAPICH2. See MPI Libraries on page 9 for details.

  — mpitests

  — mpi-selector

  — Sandia* OpenSHMEM

  — Open Fabrics Interface (OFI) libfabric

The IFS package includes the Basic package plus:

- Fabric Manager, which allows comprehensive control of administrative functions using a mature Subnet Manager. Fabric Manager simplifies subnet, fabric, and individual component management, easing the deployment and optimization of large fabrics.

- FastFabric Toolset, which enables rapid, error-free installation and configuration of Intel® OPA host software and management software tools, as well as simplified installation, configuration, validation, and optimization of HPC fabrics. For details, refer to the *Intel® Omni-Path Fabric Suite FastFabric User Guide*.

1.9 Firmware Files

This release of the Intel® Omni-Path Software contains the firmware files listed in the following table.

Note: The version number is shown in bold.

Table 1. Firmware Files

File	OS	Non-CUDA/ CUDA	RPM Version
hfi1 driver	RHEL* 7.4	Non-CUDA	kmod-ifs-kernel-updates-3.10.0_693.21.1.el7.x86_64-**1618**.x86_64.rpm
		CUDA	kmod-ifs-kernel-updates-3.10.0_693.21.1.el7.x86_64-**1618cuda**.x86_64.rpm
	RHEL* 7.5	Non-CUDA	kmod-ifs-kernel-updates-3.10.0_862.el7.x86_64-**1618**.x86_64.rpm
		CUDA	kmod-ifs-kernel-updates-3.10.0_862.el7.x86_64-**1618cuda**.x86_64.rpm
	SLES* 12.2	Non-CUDA	ifs-kernel-updates-kmp-default-4.4.114_92.67_default_k4.4.114_92.67-**1618**.x86_64.rpm
		CUDA	ifs-kernel-updates-kmp-default-4.4.114_92.67_default_k4.4.114_92.67-**1618cuda**.x86_64.rpm
	SLES* 12.3	Non-CUDA	ifs-kernel-updates-kmp-default-4.4.114_94.14_default_k4.4.114_94.14-**1618**.x86_64.rpm
		CUDA	ifs-kernel-updates-kmp-default-4.4.114_94.14_default_k4.4.114_94.14-**1618cuda**.x86_64.rpm
	SLES* 15	Non-CUDA	ifs-kernel-updates-kmp-default-4.12.14_15_default_k4.12.14_15-**1618**.x86_64.rpm

1.10 Release Compatibility

This release is backwards compatible with the most recent minor release version. For example, Release 10.N is backwards compatible with Release 10.N-1 and Release 10.N-1.x.

1.11 Operating Systems

This release of the Intel® Omni-Path Software supports the operating systems listed in the following table.

Table 2.   Supported Operating Systems

Operating System	Update/ SP	Kernel Version
Red Hat* Enterprise Linux* (RHEL*) 7.4 X86_64	Update 4	3.10.0-693.el7.x86_64
Red Hat* Enterprise Linux* (RHEL*) 7.5 X86_64	Update 5	3.10.0-862.el7.x86_64
CentOS* 7.4 X86_64	Update 4	3.10.0-693.el7.x86_64
CentOS* 7.5 X86_64	Update 5	3.10.0-862.el7.x86_64
Scientific Linux* 7.4 X86_64	Update 4	3.10.0-693.el7.x86_64
Scientific Linux* 7.5 X86_64	Update 5	3.10.0-862.el7.x86_64
SUSE* Linux* Enterprise Server (SLES*) 12.2 X86_64	Service Pack 2	4.4.21-69.1-default
SUSE* Linux* Enterprise Server (SLES*) 12.3 X86_64	Service Pack 3	4.4.73-5.1-default
SUSE* Linux* Enterprise Server (SLES*) 15 X86_64		4.12.14-23-default

Table 3.   Supported OS by CPU

CPU Type	RHEL*		SLES*			CentOS*		Scientific Linux*	
	7.4	7.5	12.2	12.3	15	7.4	7.5	7.4	7.5
Intel® Xeon Phi™ x200 Product Family (Knights Landing CPU-based servers)	•	•	•	•	•	•	•	•	•
Intel® Xeon® Scalable Processors (Skylake CPU-based servers)	•	•	•	•	•	•	•	•	•
Intel® Xeon Phi™ Processor (Knights Mill CPU-based servers)	•	•		•	•				
Next Generation Intel® Xeon® Scalable Processors (codename Cascade Lake)	•	•	•	•	•	•	•	•	•

Notes: PSM2 GPUDirect* RDMA with CUDA is supported as shown below:

- CUDA Toolkit 9.1 is supported on RHEL* 7.4, RHEL* 7.5, and SLES* 12 SP3.
- CUDA Toolkit 8.x is supported with minimal validation for SLES* 12 SP2.
- Intel does not officially support CUDA with SLES* 15 at this time.

1.12 Parallel File Systems

The following parallel file systems have been tested with this release of the Intel® Omni-Path Software:

- Lustre* Long Term Support (LTS) release, version 2.10.4
  - — on RHEL* 7.5 and SLES* 12 SP3 (via LNET Self Test)
- IBM* Spectrum Scale[1], version 5.0.1.1
  - — on RHEL* 7.4, RHEL* 7.5, SLES* 12 SP2, and SLES* 12 SP3

Note: SLES* 15 is not supported for IBM* Spectrum Scale and Lustre* in this release.

Refer to the *Intel® Omni-Path Fabric Performance Tuning User Guide* for details on optimizing parallel file system performance with Intel® Omni-Path Software.

1.13   MPI Libraries

In this release, MPI builds have been reduced to the libraries as listed in Table 4.

Note: The sources for Open MPI and MVAPICH2 are included in the software package. FastFabric provides a simple tool to build these MPIs using the Intel® Compiler (not included in the software package). Refer to *Intel® Omni-Path Fabric Suite FastFabric User Guide*, Rebuilding MPI Library and Tools.

1.13.1   Supported MPI Libraries

The table below lists the different MPI libraries supported by Intel® Omni-Path Fabric Software with the corresponding version, fabric support, and compiler used. Note that the second column indicates whether the MPI library is included in the Intel® Omni-Path Software package or not.

Table 4.   Supported MPI Libraries

MPI Implementation	Included in Basic Package?	Runs Over	Compiled With
Open MPI 2.1.2	Yes	PSM2, OFI	GCC
Open MPI 2.1.2-cuda	Yes	PSM2	GCC
MVAPICH2-2.3B	Yes	PSM2	GCC
Intel® MPI Library 2018 Update 3	No	PSM2 (via TMI), OFI	N/A
IBM* Spectrum* MPI version 10.1.1.0	No	PSM2	N/A

1.13.2   Compiler Versions and Distributions

The MPI libraries listed in the preceding section that are included in the release and built with PSM2 support were built with the following compiler versions:

Table 5.   Compiler Versions and Distributions

Compiler	OS Distribution	Compiler Version
(GNU) gcc	RHEL* 7.4	gcc version 4.8.5 20150623 (Red Hat* 4.8.5-16) (GCC)
(GNU) gcc	RHEL* 7.5	gcc version 4.8.5 20150623 (Red Hat* 4.8.5-28) (GCC)
		continued...

---

1 Formerly known as General Parallel File System (GPFS)

Compiler	OS Distribution	Compiler Version
(GNU) gcc	SLES* 12 SP 2	gcc version 4.8.5 (SUSE* Linux*)
(GNU) gcc	SLES* 12 SP 3	gcc version 4.8.5 (SUSE* Linux*)
(GNU) gcc	SLES* 15	gcc version 7.3.1 20180307 (SUSE* Linux*)

Note: Refer to the *Intel® Omni-Path Fabric Host Software User Guide* for setup information when using Open MPI with the SLURM PMI launcher and PSM2.

1.14 Intel Hardware

The following table lists the Intel hardware supported in this release. The table does not include OEM-specific hardware, such as custom adapters and switches.

Note: The Intel® PSM2 implementation has a limit of four (4) HFIs.

Table 6. Supported Hardware

Hardware	Description
Intel® Xeon® Processor E5-2600 v3 product family	Haswell CPU-based servers
Intel® Xeon® Processor E5-2600 v4 product family	Broadwell CPU-based servers
Intel® Xeon® Scalable Processors	Skylake CPU-based servers
Next Generation Intel® Xeon® Scalable Processors	Cascade Lake CPU-based servers
Intel® Xeon Phi™ x200 Product Family	Knights Landing CPU-based servers
Intel® Xeon Phi™ 72x5 Processor Family	Knights Mill CPU-based servers
Intel® Omni-Path Host Fabric Interface 100HFA016 (x16)	Single Port Host Fabric Interface (HFI)
Intel® Omni-Path Host Fabric Interface 100HFA018 (x8)	Single Port Host Fabric Interface (HFI)
Intel® Omni-Path Switch 100SWE48Q	Managed 48-port Edge Switch
Intel® Omni-Path Switch 100SWE48U	Externally-managed 48-port Edge Switch
Intel® Omni-Path Switch 100SWE48UFH	Externally-managed 48-port Edge Switch, hot-swap power and fans
Intel® Omni-Path Switch 100SWE48QFH	Managed 48-port Edge Switch, hot-swap power and fans
Intel® Omni-Path Switch 100SWE24Q	Managed 24-port Edge Switch
Intel® Omni-Path Switch 100SWE24U	Externally-managed 24-port Edge Switch
Intel® Omni-Path Director Class Switch 100SWD24	Director Class Switch 100 Series, up to 768 ports
Intel® Omni-Path Director Class Switch 100SWD06	Director Class Switch 100 Series, up to 192 ports

1.15 Switch Firmware

The following firmware is supported for Intel® Omni-Path switches:

- Intel® Omni-Path Switch Firmware 10.8.x revision (managed and externally-managed switches)
- Intel® Omni-Path Switch Firmware 10.7.x revision (managed and externally-managed switches)

Refer to the *Intel® Omni-Path Fabric Switches Release Notes* for more information.

1.16 Document Versions

The following table lists the end user document versions supported by this release.

Table 7. **Supported Document Versions**

Title	Doc. Number	Revision
Intel® Omni-Path Fabric Quick Start Guide	J57479	5.0
Intel® Omni-Path Fabric Setup Guide	J27600	9.0
Intel® Omni-Path Fabric Switches Hardware Installation Guide	H76456	7.0
Intel® Omni-Path Host Fabric Interface Installation Guide	H76466	5.0
Intel® Omni-Path Fabric Software Installation Guide	H76467	10.0
Intel® Omni-Path Fabric Switches GUI User Guide	H76457	10.0
Intel® Omni-Path Fabric Switches Command Line Interface Reference Guide	H76458	10.0
Intel® Omni-Path Fabric Suite FastFabric User Guide	H76469	10.0
Intel® Omni-Path Fabric Suite Fabric Manager User Guide	H76468	10.0
Intel® Omni-Path Fabric Suite Fabric Manager GUI User Guide	H76471	10.0
Intel® Omni-Path Fabric Host Software User Guide	H76470	10.0
Intel® Performance Scaled Messaging 2 (PSM2) Programmer's Guide	H76473	10.0
Intel® Omni-Path Fabric Performance Tuning User Guide	H93143	12.0
Intel® Omni-Path IP and LNet Router Design Guide (Old title: Intel® Omni-Path IP and Storage Router Design Guide)	H99668	7.0
Building Containers for Intel® Omni-Path Fabrics using Docker* and Singularity* Application Note	J57474	5.0
Intel® Omni-Path Management API Programmer's Guide	J68876	4.0
Configuring Non-Volatile Memory Express* (NVMe*) over Fabrics on Intel® Omni-Path Architecture Application Note	J78967	1.0
Intel® Omni-Path Fabric Software Release Notes	K21143	1.0
Intel® Omni-Path Fabric Manager GUI Release Notes	K21144	1.0
Intel® Omni-Path Fabric Switches Release Notes (includes managed and externally-managed switches)	K21142	1.0
Intel® Omni-Path Fabric Unified Extensible Firmware Interface (UEFI) Release Notes	K21145	1.0
Intel® Omni-Path Fabric Thermal Management Microchip (TMM) Release Notes	K21147	1.0
Intel® Omni-Path Fabric Firmware Tools Release Notes	K21148	1.0

Related Links

1.17 Installation Requirements

1.17.1　Software and Firmware Requirements

Table 2 on page 8 lists the operating systems supported by this release. Refer to the *Intel® Omni-Path Fabric Software Installation Guide* for the required packages.

1.17.2　Installation Instructions

There are two Intel® Omni-Path Fabric Software packages:

- `IntelOPA-IFS.<distro>-x86_64.<version>.tgz` for the management node.
- `IntelOPA-Basic.<distro>-x86_64.<version>.tgz` for compute nodes.

The packages in the tgz file are RPMs. Installing individual RPMs is not supported in this release.

Refer to the *Intel® Omni-Path Fabric Software Installation Guide* for related software requirements and complete installation procedures. Refer to the *Intel® Omni-Path Fabric Switches Hardware Installation Guide* for related firmware requirements.

1.17.3　Installation Path Changes in Release 10.4 (and later)

If you are upgrading an Intel® Omni-Path Fabric Software installation (Release 10.3 or earlier), Intel recommends that you perform the following steps before upgrading, due to changes in installation paths for RPMs and configuration files.

Run `./INSTALL -u` to uninstall existing packages.

Run `./INSTALL -a` to complete the installation.

Pre-existing configuration files are automatically saved by the RPM as `.rpmsave` files. (RPM will notify you about these files during removal.) If you want to keep these configuration files, you should move them to their new locations. A mapping of old configuration file locations to new locations is shown in the following table.

Old Location (Release 10.3 and earlier)	New Location (Release 10.4 and later)
`/etc/sysconfig/opafm.xml`	`/etc/opa-fm/opafm.xml`
`/etc/sysconfig/allhosts`	`/etc/opa/allhosts`
`/etc/sysconfig/chassis`	`/etc/opa/chassis`
`/etc/sysconfig/esm_chassis`	`/etc/opa/esm_chassis`
`/etc/sysconfig/hosts`	`/etc/opa/hosts`
`/etc/sysconfig/opafastfabric.conf`	`/etc/opa/opafastfabric.conf`
`/etc/sysconfig/opaff.xml`	`/etc/opa/opaff.xml` Changed to `/etc/opa/opamgt_tls.xml` in 10.7.
`/etc/sysconfig/opamon.conf`	`/etc/opa/opamon.conf`
`/etc/sysconfig/ports`	`/etc/opa/ports`
`/etc/sysconfig/switches`	`/etc/opa/switches`

1.18 Product Constraints

- Power class 2 AOC are supported. You must use 10.5 (or newer) host software and 1.5 (or newer) UEFI for proper operation. Integrated HFI (-F) requires a specific BIOS level to support power class 2 AOC; contact your BIOS vendor for more information.

- On certain systems with more than 1 HFI, the ports can come up in an unexpected way. This can lead to the wrong HFI being associated with ib0 or ib1. To ensure ports come up in the expected order each time, use the module parameter `hfi1 port_reorder=1` and if you are loading the driver during initramfs boot, then ensure that you rebuild initramfs.

- Intel® MPI as a component of Intel® Parallel Studio 2018, Update 3 does not support SLES* 15.

1.19 Product Limitations

This release has the following product limitations:

- The embedded version of the Fabric Manager supports a maximum of 100 HFI ports involving less than 20 switch ASICs. Calculate the number of switch ASICs in your fabric as follows:

  — One ASIC per Intel® Omni-Path Edge Switch 100 Series

  — Two ASICs per Intel® Omni-Path Director Class Switch 100 Series Leaf module

  — Two ASICs per Intel® Omni-Path Director Class Switch 100 Series Spine module

- Performance Administration (PA) Failover should **not** be enabled with FMs running on differing software versions.

  To disable PA failover, edit the `/etc/sysconfig/opafm.xml` file and in the `<Pm>` section, change `<ImageUpdateInterval>` to `0`.

- Enabling UEFI Optimized Boot on some platforms can prevent the HFI UEFI driver from loading during boot. To prevent this, do not enable UEFI Optimized Boot.

1.20 Accelerated RDMA Information

Accelerated RDMA is a Verbs protocol extension to improve the performance of RDMA write and RDMA read operations on Intel® Omni-Path hardware.

This extension improves the efficiency of large message transfers to provide performance benefits for storage protocols and other Verbs-based protocols. The benefits include increased achievable bandwidth with reduced CPU utilization. The Token ID (TID) RDMA protocol accelerates the OpenFabrics Alliance* (OFA) Verbs API with no changes required to API consumers. The acceleration technique is performed by the host driver and the application running over the OFA Verbs API does not need to make any code change.

Accelerated RDMA is off by default.

To enable it, add `cap_mask=0x4c09a01cbba` to the `/etc/modprobe.d/hfi1.conf` file. Instructions on how to do this are in the *Intel® Omni-Path Fabric Performance Tuning User Guide*, Setting HFI1 Driver Parameters section.

Note: It is not necessary to enable Accelerated RDMA on all nodes in a cluster. The performance effects of enabling Accelerated RDMA on a subset of nodes has not been characterized. It is highly recommended that you test to see if enabling this feature works best for your applications.

2.0   Issues

This section lists the resolved and open issues in the Intel® Omni-Path Software.

2.1   Resolved Issues

2.1.1   Issues Resolved in this Release

The following table lists issues that are resolved in this release.

Table 8.   Issues Resolved in this Release

ID	Description	Resolved in Release
STL-46790	In cases where GSI services are active and the FM is receiving capability change traps (common after node reboots), FM responsiveness may be impacted. This could result in data traffic disruption or unexpected FM failovers. GSI traffic would include the PM, SA, and DBSync (FM failover).	10.8.0.2
STL-46606 STL-47956 STL-48661	Bouncing a link or rebooting a device under certain fabric conditions may cause a switch in the fabric to be removed from the Fabric Manager's internal view of the topology leading to fabric disruptions and instability.	10.8.0.1
STL-47546	When an ISL goes down in the middle of an FM sweep (due to a disruption in the fabric such as a reboot), the SA copy of topology becomes invalid when the Fattree routing algorithm is used. SA queries that use this topology (e.g., path record query) fail. *Note:* A path record query failure can be seen in FM log as "INVALID TOPOLOGY" messages. The issue will resolve after the FM's next successful sweep.	10.8.0.1
135981 143631	In RHEL 7.3, irqbalance spams messages every ten seconds with snippet below. The HFI driver sets affinity and the Omni Fabric install sets `/etc/sysconfg/irqbalance` to use exact. `IRQBALANCE_ARGS= --hintpolicy=exact` Irq balancer scans the `isolcpus` and `nohz_full` kernel masks and adds the corresponding CPUs to the banned_cpus mask. This works fine for valid masks, but not for the default, empty masks. In this case, when they read from the `sysfs` they return empty strings, "\n" or "0x0, \n" *Note:* RHEL* 7.3 is no longer supported.	10.8
136432	Certain perftest tools such as ib_write_bw do not work on RHEL* 7.3 when using the RDMA CM with UD QPs. *Note:* RHEL* 7.3 is no longer supported.	10.8
140797 141558 141852 142476	When installing an IFS tar package on a server where the OS-resident OPA software has been installed, the installation may fail to properly update all the RPMs. This is due to the use of Epoch tagging on the OS-resident OPA software which makes the OS-resident RPMs appear to be newer than RPMs in the IFS tar package.	10.8
141263	The GUPS application that is included as part of Sandia Open Shmem apps may fail with more than 1% of error when executed with more that four PEs. Note: GUPs does not use atomic XOR operations given they are not supported in the OpenShmem specification, version 1.3. This is addressed in version 1.4.	10.8

continued...

ID	Description	Resolved in Release
141636	During the Intel OPA installation, the ramdisk may be rebuilt multiple times; however, only the last build is necessary.	10.8
141586	Using SLES* 12 SP2 and SP3, INSTALL -a fails due to dependency issues. The following information to resolve this issue has been added to the *Intel® Omni-Path Fabric Software Installation Guide*, OS RPMs Installation Prerequisites section: "Prior to installing Intel OPA, install the SLES SDK iso from SUSE which contains additional, required development packages."	10.8
141793	Use of a static buffer could produce an incorrect device name (hfi1_x) in dmesg logging.	10.8
142330	MPI applications that leverage the PSM2 library's access to the HFI ASICs Memory Mapped IO and that access the MMIO directly (not via PSM2) can potentially cause an "unsupported opcode" error which some servers handle as a critical error. *Note:* A side effect of this issue is that the hfi1 driver automatically masks the AER Unsupported Request error for the OPA1 HFI device.	10.8
143311	During UEFI pre-boot with the connected switch running 10.7.0.0.134, the OPA link may not complete initialization if the link is bounced or restarted. This behavior is limited to the pre-boot period. There is no exposure once Linux boot has completed.	10.8
143449	PM will scroll LQI=0 and Integrity Exceeded Threshold logs when an additional VF with QoS enabled and a device group that is not "All". *Note:* This issue does not occur when running against the default opafm.xml configuration file.	10.8
143915	When specific messages that are of non-double word length (not 4 B) and less than 1 MTU in size (10,240 B in most configurations) are sent, they may never complete, causing an application to wait indefinitely. *Note:* This issue occurs on machines with more than 28 CPU cores (not including hyper-threads) or more than 28 HFI contexts enabled per Intel® Omni-Path HFI.	10.8
144165	Nodes unable to ping on IPoIB. *Note:* This issue occurs when a host port disappears and reappears from the FM's topology (usually due discovery timeout or major fabric disruption), while the port remains ACTIVE the entire time. This results in the host port not being a member of the IP multicast groups. The primary symptom is the inability to resolve IP addresses via ARP.	10.8
144795	ESM assigns invalid MLIDs. IPoIB and other applications that use multicast will not function correctly with invalid MLIDs.	10.8
145399 146088	Using the `./INSTALL -i` command results in errors due to missing dependencies.	10.8
145917	As required by the System Management Bus (SMBus) specification, when the Bus is idle for less than 50 µs, the HFI card pulls the Bus into low voltage. However, the Bus does not remain in low voltage for a long enough interval.	10.8

2.1.2   Issues Resolved in Prior Releases

The following table lists issues that were resolved in prior releases.

Table 9.   Issues Resolved in Prior Releases

ID	Description	Resolved in Release
131745	When running Open MPI 1.10.0 on SLES* 12 with large number of ranks per node (over 40), it may happen that the ORTE daemon (orted) "hangs" during the finalization of job. This is an issue in Open MPI with the version of glibc used in SLES* 12. It is being researched by the Open MPI community in issue: https://github.com/open- mpi/ompi/issues/1136	10.7
134471	The HFI UEFI driver cannot boot via PXE using Grub 2.	10.7
		continued...

ID	Description	Resolved in Release
134494	Open MPI uses srand() family functions at MPI_Init() time. Therefore, if the user sets srand() before calling MPI_Init(), the values will be altered.	10.7
135084	In rare circumstances, the HFI may not appear in the PCI config space after a power cycle.	10.7
135975	After performing an OPA software configuration update, switches will show the new settings when queried by opaswitchadmin tools, however, individual ports will continue to operate using the previous settings, including LinkWidth enable.	10.7
136160	On some Intel® Xeon Phi™ with integrated Intel® Omni-Path fabric platforms, the second integrated HFI is discovered first and is subsequently identified as the first HFI device. As a result, when issuing Intel® Omni-Path commands, the second HFI appears first in the results. In Linux* and various Intel® Omni-Path tools, the HFI reporting order may be the opposite of the order appearing on the Intel® Xeon Phi™ with integrated Intel® Omni-Path fabric cable/faceplate.	10.7
136500	RDMA perftests can hang on start on a client side when RDMA CM (-R option) is used.	10.7
136728	If hundreds of links are bouncing while the FM is sweeping, the FM sweep time may be significantly extended. This can result in unexpected delays in FM responsiveness to fabric changes or host reboots. (The issue is that active links bounce between the time FM discovers one side of the link versus the other side of the link.) In Release 10.3.1, a change was made to improve the FM responsiveness in large fabrics of >1000 nodes when numerous links bounce (or hosts are rebooted) at once. In Release 10.6, a new configuration value is present in the FM configuration that determines how much time will be allotted to timeouts before abandoning a sweep. If you are upgrading from a previous version of the FM and retaining a configuration file that does not include this new parameter, the value will be set too low and cause sweeps to abandon after only a single timeout is witnessed.	10.7
137951	In the HFI BIOS screen for Advanced NIC Configuration, a warning message about incorrect custom P_Key value is not completely displayed.	10.7
139550	Infrequently, an AOC may exhibit an unexpectedly high local link integrity error rate after the link comes up, relative to the error rate on previous link up occasions. This can be determined by observing a link quality of <5. These links may eventually experience a link width downgrade.	10.7
139981	When Accelerated RDMA (TID RDMA) is enabled, GPFS file system can hang.	10.7
140691	When running `opaswitchadmin` against multiple externally-managed switches simultaneously, it sends schedule requests in parallel to those hosts. It is possible that some hosts may intermittently fail due to timeouts at high levels of parallelism.	10.7
140707	When using the Upgrade option in the INSTALL script TUI, a message is displayed stating that "Up To Date" items will be reinstalled. However, the opa.log file does not indicate these items were reinstalled.	10.7
140881	In rare cases when an LNI failure occurs, the link will not come up after manually disabling and re-enabling the link.	10.7
140911	In Release 10.6, the OFI verbs provider does not support FI_EP_RDM End Point type. This End Point type is needed for Open MPI OFI support. Therefore, Open MPI OFI support will not run over the verbs provider.	10.7

141219	When adaptive routing is disabled, the output for `opasmaquery` for `portgroup` appears as shown below:
	
	`# opasmaquery -l 1 -o portgroup`
	`PG: 0x0000 Egress:None`
		10.7

141420	When Accelerated RDMA is enabled, the kdeth_qp module parameter cannot be changed.	10.7
141558 141852	In IFS versions 10.3.1 and earlier, the rpm contains the Epoch tag. When upgrading to a newer version without the Epoch tag (i.e., 10.4 through 10.6.x), the rpm tools act as though the old rpm is the newer version. This issue causes a failure in the IFS upgrade.	10.7

continued...

ID	Description	Resolved in Release
141782	In Release 10.7, the `opapmaquery -n` argument no longer accepts decimal input. In the user documentation and the man pages, the `-n` argument is described as the port in hexadecimal. Until this release, the tool also accepted decimal input.	10.7
141845	Resolved FM process out of memory condition	10.7
141909	Resolved multiple FM synchronization issue that can lead to FM failure.	10.7

2.2 Open Issues

The following table lists the open issues for this release.

Table 10. Open Issues

ID	Description	Workaround
129563	Memory allocation errors with MVAPICH2-2.1/Verbs.	*Note:* To avoid this issue, use MPIs over PSM. If you are using MPIs over verbs, the following workaround is required: • When running MVAPICH2 jobs with a large number of ranks (for example, > 36 ranks but ≤ 72 ranks), you must set the following parameters in `/etc/security/limits.conf`: — hard memlock unlimited — soft memlock unlimited • Also, you must increase the lkey_table_size:LKEY table size in bits (2^n, where $1 \le n \le 23$) from its default of 16 to 17. For instructions on setting module parameters, refer to the *Intel® Omni-Path Fabric Performance Tuning User Guide*, HFI1 Driver Module Parameters chapter.
132207	Kernel crash caused by the ib_srpt module.	Install this kernel patch: https://git.kernel.org/cgit/linux/kernel/git/torvalds/linux.gi t/commit/?id=51093254bf879bc9ce96590400a87897c749 8463
133604	Bonding driver shows incorrect hardware address of IPoIB interfaces.	This workaround applies to SLES* 12 SP3 (and earlier) distributions: Use the `opainfo` command to retrieve the PortGUID and `ip addr show ib0` to get the correct 20-byte hardware address of OPA network interface.
134819	In KNL-F EFI shell, the command `ifconfig -l` does not correctly display the IP address after being assigned via DHCP.	Launch a newer version of the EFI shell from the embedded shell.
135028	NVMe over Fabric Protocol is only supported on Intel® OPA with Linux* kernel 4.5 and later versions.	To use NVMe functionality on Intel® OPA, you must patch the kernel.
135830	On Intel® Xeon Phi™ systems, failure observed during software upgrade when rebuilding the boot image. Error message contains: Rebuilding boot image with "/usr/bin/dracut -f"	Install parallel gzip (pigz) and make it the default compression tool prior to upgrading the software.
136822	The Intel UEFI driver contained in the server BIOS must be executed for proper support of Active Optical Cables (AOC) in an integrated HFI environment. Some BIOS do not execute the UEFI in Legacy BIOS Boot mode, and there are BIOS configuration settings that may prevent the UEFI from executing in any mode.	Avoid the use of Legacy BIOS boot mode if your platform does not execute the HFI driver in that mode. Avoid BIOS settings or other configuration settings that do not execute the HFI driver during boot.

continued...

ID	Description	Workaround
139368	Some applications compiled with older compilers may use a personality bit that signifies that READ should imply EXECUTE permissions. To improve system security, the hfi1 driver does not allow execute permissions on PSM memory maps. Therefore, applications that use READ implies EXECUTE will fail to run.	As root, run the execstack tool to clear the executable bit on the binary: `execstack –c <binary>` Alternatively, recompile the binary to not set this personality bit.
139613	The Subsystem Vendor and Subsystem Device ID in the PCI configuration space of Intel® Omni-Path discrete HFI cards may not indicate the correct OEM vendor and device. As a result, the `lspci` command may show incorrect Subsystem Vendor and Device ID information. This issue affects Intel server boards for Intel® Xeon® Processor v3 and v4 Product Family configured in Legacy OS boot mode.	Reconfigure the system from Legacy OS boot mode to UEFI boot mode.
139743 143031 143115	Under a very heavy load through the IPoIB interface, the kernel warning `NETDEV WATCHDOG: ib0 (hfi1): transmit queue 0 timed out`, followed by the messages `queue stopped 1, tx_head xxx, tx_tail xxx` and `transmit timeout: latency xxxx msecs` may be seen.	Reduce traffic pressure on ipoib interface to resume transmission.
139924	For RHEL* 7.4 (and earlier) and SLES*, the ibacmp provided in the OS distribution uses incorrect address information when joining multicast groups. This causes name resolution to fail. The dsap provided in the OS distribution works correctly.	The fix for this issue is available in the library rdma-core-15-2.el7. • Do not use the OS distribution ibacmp. • Install rdma-core-15-2.el7 manually.
139995	When installing Intel® Omni-Path Software on a clean system (with no prior versions installed), the following error message may be displayed: `cat: //etc/opa/version_delta: No such file or directory`	This message can be safely ignored. The installer is looking for an IFS version file before it has been created on the system. The installation is not impacted.
140310	On RHEL* 7.4 and RHEL* 7.5, if an ipofabric interface name is non- standard, the ifcfg file is not read correctly. For example, if you are using the ipofabric interface name opa_ib0, and the connected mode and MTU size is specified in the ifcfg-opa_ib0 file, when you bring up the interface with ifup, the settings do not take effect.	You must manually enter: `echo connected > /sys/class/net/opa_ib0/mode` for the settings to take effect.
141005	In 10.8, the output of the module parameter `num_user_contexts` has been changed. The value of the module parameter `num_user_context` used at driver init time determines the number of receive contexts reserved for PSM. The default value of -1 caused the driver init sequence to determine how many CPUs are available, and assigned the context count to that value. *It would then update the module parameter `num_user_contexts` with that value.* This incorrect behavior has been fixed. `num_user_contexts` is no longer updated.	Do not use `num_user_contexts` to determine how many context can be used. To determine the number of available user contexts, refer to the sysfs file: `/sys/class/infiniband/hfi1_<port>/nctxts`
141273	The in-distro version of perftests has bugs.	Use the upstream version of perftest from https://github.com/linux-rdma/perftest.
141740	Due to the race condition during boot up ipoib driver can miss PKey change event that will leave ipoib interface in disconnected state.	Reload ipoib driver on the affected host: `#> modprobe –r ib_ipoib && modprobe ib_ipoib`

continued...

ID	Description	Workaround
143031	Under a very heavy load through the IPoIB interface, the kernel warning NETDEV WATCHDOG: ib0 (hfi1): transmit queue 0 timed out, followed by the messages queue stopped 1, tx_head xxx, tx_tail xxx and transmit timeout: latency xxxx msecs may be seen.	The default ib_ipoib module option `send_queue_size` is 256. To avoid TX timeouts, Intel recommends that you modify `send_queue_size` to 2048 or higher, up to a maximum of 8192.
143174	Due to a SLES* 15 kernel setting, hfidiags cannot work while the HFI driver is loaded. The tool and driver are mutually exclusive.	Boot the kernel with `iomem=relaxed`. This turns off the resource exclusivity check.
143296	When irqbalance uses the argument `--hintpolicy=exact`, it applies the policy of setting the hardware interrupts to CPU core mappings according to device drivers preferences. For the HFI1 driver, it is strongly recommended to preserve interrupt locality for low latency and high bandwidth by having a dedicated CPU core per interrupt.	Always start the user-space process irqbalance using the argument `--hintpolicy=exact`.
143915	When specific messages that are of non-double word length (not 4 B) and less than 1 MTU in size (10,240 B in most configurations) are sent, they may never complete, causing an application to wait indefinitely. *Note:* This issue occurs on machines with more than 28 CPU cores (not including hyper-threads) or more than 28 HFI contexts enabled per Intel® Omni-Path HFI.	Add this command line option for Open MPI: `-x PSM2_CUDA_THRESH_RNDV=8192`
144996	Running workloads with more than 78 ranks with the Open MPI OFI MTL over OFI Verbs;OFI_RXM provider may result in a hang with message sizes larger than 65 KB.	To avoid this issue, increase the TX/RX size by setting the following environment variables on the mpirun command line: • `-x FI_VERBS_RX_SIZE=2000` • `-x FI_VERBS_TX_SIZE=2000`
145415	ESM hangs after the first sweep in a configuration that includes all `<VirtualFabric>` sections that are enabled and with `QOS =1`, where `<BaseSL>` is defined in all with values other than `0`. Symptoms of this hang are: • SA or PA queries to ESM time out (e.g., opareport fails with FTIMEOUT) • SM does not resweep after SweepInterval, multicast membership change, or fabric change • smControl shutdown is issued but ESM never reaches stopped state	Ensure that the FM is configured where `<VirtualFabric>` sections are enabled, have `QOS=0` or if `QOS=1`, and at least one VF has `BaseSL=0`.
145474	OFI Verbs mpi_stress may cause verbs/MSG provider completion queue overrun that results in dropped completions. They show up as sequence errors in the test.	When using Intel MPI, ensure that the transmit and receive queue sizes are large enough for the given number of ranks and expected number of messages that would be sent/received. The corresponding environment variables must be increased to avoid errors. The verbs/MSG provider CQ size is internally calculated as `(FI_OFI_RXM_MSG_TX_SIZE + FI_OFI_RXM_MSG_RX_SIZE) * FI_UNIVERSE_SIZE`
145585	SLES* 15 `./INSTALL` script does not properly run `dracut -f`.	Run `dracut -f` after the IFS installation.
145623	For systems running on SLES* 15, there is a known issue with irqbalance.	Contact Intel Customer Support for more information.

continued...

ID	Description	Workaround
145771	Due to changes in the SLES* 15 kernel, a user space application cannot access a hardware resource if that resource is being used by a kernel driver. The result is that the hfi1_eprom cannot access the EEPROM on an HFI when hfi1 kernel driver is using the device.	Perform the following workaround: 1. Unload hfi1 kernel driver: `rmmod hfi1` 2. Run hfi1_eprom commands. 3. Reload hfi1 kernel driver: `modprobe hfi1`
145855	If the Admin VF is not running on VL0, the HSM may get into a state where it is unable to talk to the fabric. The sweep will log the following errors: ```opamgt ERROR: [<pid>] omgt_send_mad2: send failed; Invalid argument, agent id 2 MClass 0x81 method 0x1 attrId 0x11 attrM 0x0 WARN [topology]: SM: sm_send_stl_request_impl: Error Sending to Path:[1] Lid:[0xffffffff] [Can't find node in topology!]. AID:[NODEINFO] TID:[0x0000000000000031] Status:[OK (0x00000000)] WARN [topology]: SM: topology_main: TT: too many errors during sweep, will re-sweep in a few seconds rc: 108: unrecoverable error```	Perform the following: • Ensure the FM configuration defines the Admin VF before any other VFs. • If you are using the V10.8 Multi-tenancy feature, ensure the QOSGroup used by the Admin VF is the first QOSGroup defined in the `<QOSGroups>` section of the FM configuration. *Note:* The Admin VF may be configured to run on any SL. As long as the Admin VF is defined first, the SM will map the Admin VF's SL to VL0.
146271	When using `libpsm2` with `PSM2_CUDA=1`, workloads may assert() during MPI_Recv or MPI_ collective operations where the source buffer is a GPU buffer and located on the same node, i.e., intranode transfer. The destination buffer locality has no influence. This assert() will only affect transfers that are not aligned to the start of the source buffer.	Download, build, and install the latest libpsm2 located at https://github.com/intel/opa-psm2 *Note:* The fix will be made available shortly AFTER the release of libspm2 v11.2.23. Please refer to the release notes on github to determine the proper version to install.

3.0　Related Information

3.1　Intel® Omni-Path Documentation Library

Intel® Omni-Path publications are available at the following URLs:

- Intel® Omni-Path Switches Installation, User, and Reference Guides

  http://www.intel.com/omnipath/SwitchPublications

- Intel® Omni-Path Software Installation, User, and Reference Guides (includes HFI documents)

  http://www.intel.com/omnipath/FabricSoftwarePublications

- Drivers and Software (including Release Notes)

  http://www.intel.com/omnipath/Downloads

Use the tasks listed in this table to find the corresponding Intel® Omni-Path document.

Task	Document Title	Description
Key: Shading indicates the URL to use for accessing the particular document.		
• Intel® Omni-Path Switches Installation, User, and Reference Guides: http://www.intel.com/omnipath/SwitchPublications		
• Intel® Omni-Path Software Installation, User, and Reference Guides (includes HFI documents): http://www.intel.com/omnipath/FabricSoftwarePublications (no shading)		
• Drivers, Software, and Firmware (including Release Notes): http://www.intel.com/omnipath/Downloads		
Using the Intel® OPA documentation set	*Intel® Omni-Path Fabric Quick Start Guide*	A roadmap to Intel's comprehensive library of publications describing all aspects of the product family. This document outlines the most basic steps for getting your Intel® Omni-Path Architecture (Intel® OPA) cluster installed and operational.
Setting up an Intel® OPA cluster	*Intel® Omni-Path Fabric Setup Guide*	Provides a high level overview of the steps required to stage a customer-based installation of the Intel® Omni-Path Fabric. Procedures and key reference documents, such as Intel® Omni-Path user guides and installation guides, are provided to clarify the process. Additional commands and best known methods are defined to facilitate the installation process and troubleshooting.
Installing hardware	*Intel® Omni-Path Fabric Switches Hardware Installation Guide*	Describes the hardware installation and initial configuration tasks for the Intel® Omni-Path Switches 100 Series. This includes: Intel® Omni-Path Edge Switches 100 Series, 24 and 48-port configurable Edge switches, and Intel® Omni-Path Director Class Switches 100 Series.
	Intel® Omni-Path Host Fabric Interface Installation Guide	Contains instructions for installing the HFI in an Intel® OPA cluster.
		continued...

Task	Document Title	Description
Installing host software Installing HFI firmware Installing switch firmware (externally-managed switches)	*Intel® Omni-Path Fabric Software Installation Guide*	Describes using a Text-based User Interface (TUI) to guide you through the installation process. You have the option of using command line interface (CLI) commands to perform the installation or install using the Linux* distribution software.
Managing a switch using Chassis Viewer GUI Installing switch firmware (managed switches)	*Intel® Omni-Path Fabric Switches GUI User Guide*	Describes the Intel® Omni-Path Fabric Chassis Viewer graphical user interface (GUI). This document provides task-oriented procedures for configuring and managing the Intel® Omni-Path Switch family. Help: GUI embedded help files
Managing a switch using the CLI Installing switch firmware (managed switches)	*Intel® Omni-Path Fabric Switches Command Line Interface Reference Guide*	Describes the command line interface (CLI) task information for the Intel® Omni-Path Switch family. Help: -help for each CLI
Managing a fabric using FastFabric	*Intel® Omni-Path Fabric Suite FastFabric User Guide* (Merged with: *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide*)	Provides instructions for using the set of fabric management tools designed to simplify and optimize common fabric management tasks. The management tools consist of Text-based User Interface (TUI) menus and command line interface (CLI) commands. Help: -help and man pages for each CLI. Also, all host CLI commands can be accessed as console help in the Fabric Manager GUI.
Managing a fabric using Fabric Manager	*Intel® Omni-Path Fabric Suite Fabric Manager User Guide*	The Fabric Manager uses a well defined management protocol to communicate with management agents in every Intel® Omni-Path Host Fabric Interface (HFI) and switch. Through these interfaces the Fabric Manager is able to discover, configure, and monitor the fabric.
	Intel® Omni-Path Fabric Suite Fabric Manager GUI User Guide	Provides an intuitive, scalable dashboard and set of analysis tools for graphically monitoring fabric status and configuration. This document is a user-friendly alternative to traditional command-line tools for day-to-day monitoring of fabric health. Help: Fabric Manager GUI embedded help files
Configuring and administering Intel® HFI and IPoIB driver Running MPI applications on Intel® OPA	*Intel® Omni-Path Fabric Host Software User Guide*	Describes how to set up and administer the Host Fabric Interface (HFI) after the software has been installed. The audience for this document includes cluster administrators and Message-Passing Interface (MPI) application programmers.
Writing and running middleware that uses Intel® OPA	*Intel® Performance Scaled Messaging 2 (PSM2) Programmer's Guide*	Provides a reference for programmers working with the Intel® PSM2 Application Programming Interface (API). The Performance Scaled Messaging 2 API (PSM2 API) is a low-level user-level communications interface.
Optimizing system performance	*Intel® Omni-Path Fabric Performance Tuning User Guide*	Describes BIOS settings and parameters that have been shown to ensure best performance, or make performance more consistent, on Intel® Omni-Path Architecture. If you are interested in benchmarking the performance of your system, these tips may help you obtain better performance.
Designing an IP, LNet or storage router on Intel® OPA	*Intel® Omni-Path IP and LNet Router Design Guide* (Old title: *Intel® Omni-Path IP and Storage Router Design Guide*)	Describes how to install, configure, and administer an IPoIB router solution (Linux* IP or LNet) for inter-operating between Intel® Omni-Path and a legacy InfiniBand* fabric.

continued...

Task	Document Title	Description
Building a Lustre* Server using Intel® OPA	(OBSOLETE) *Building Lustre* Servers with Intel® Omni-Path Architecture Application Note*	This document has been removed from the Intel® OPA Documentation Library. For information on how to build and configure a Lustre* server with Intel® OPA, see the Lustre* wiki: http://wiki.lustre.org.
Building Containers for Intel® OPA fabrics	*Building Containers for Intel® Omni-Path Fabrics using Docker* and Singularity* Application Note*	Provides basic information for building and running Docker* and Singularity* containers on Linux*-based computer platforms that incorporate Intel® Omni-Path networking technology.
Writing management applications that interface with Intel® OPA	*Intel® Omni-Path Management API Programmer's Guide*	Contains a reference for programmers working with the Intel® Omni-Path Architecture Management (Intel OPAMGT) Application Programming Interface (API). The Intel OPAMGT API is a C-API permitting in-band and out-of-band queries of the FM's Subnet Administrator and Performance Administrator.
Using NVMe* over Fabrics on Intel® OPA	*Configuring Non-Volatile Memory Express* (NVMe*) over Fabrics on Intel® Omni-Path Architecture Application Note*	Describes how to implement a simple Intel® Omni-Path Architecture-based point-to-point configuration with one target and one host server.
Learning about new release features, open issues, and resolved issues for a particular release	*Intel® Omni-Path Fabric Software Release Notes*	
	Intel® Omni-Path Fabric Manager GUI Release Notes	
	Intel® Omni-Path Fabric Switches Release Notes (includes managed and externally-managed switches)	
	Intel® Omni-Path Fabric Unified Extensible Firmware Interface (UEFI) Release Notes	
	Intel® Omni-Path Fabric Thermal Management Microchip (TMM) Release Notes	
	Intel® Omni-Path Fabric Firmware Tools Release Notes	

3.1.1 How to Search the Intel® Omni-Path Documentation Set

Many PDF readers, such as Adobe* Reader and Foxit* Reader, allow you to search across multiple PDFs in a folder.

Follow these steps:

1. Download and unzip all the Intel® Omni-Path PDFs into a single folder.

2. Open your PDF reader and use **CTRL-SHIFT-F** to open the Advanced Search window.

3. Select **All PDF documents in...**

4. Select **Browse for Location** in the dropdown menu and navigate to the folder containing the PDFs.

5. Enter the string you are looking for and click **Search**.

Use advanced features to further refine your search criteria. Refer to your PDF reader Help for details.